

FORM 1 (cDNA sequence provided):

1 ATGGACAGAG TTTATGAAAT TCCTGAGGAG CCAAATGTGG ATCCGGTTTC
 51 ATCTCTGGAG GAAGATGTCA TCCGTGGAGC CAACCCCGA TTTACTTTTC
 101 CATTTAGCAT CCTTTTCTCC ACCTTTTGT ACTGTGGGGA GGCTGCATCT
 151 GCTTTGTACA TGGTTAGAAT CTATCGAAAG AATAGTAAA CTTACCGGAT
 201 GACATACACC TTTTCTTCT TTATGTTTC ATCCATTATG GTCCAGTTGA
 251 CCCTCATTTT TGTCCACAGA GATCTAGCCA AAGATAAAC GCTATCATTA
 301 TTTATGCATC TAATCCTCTT GGGACCTGTT ATCAGATGTT TGGAGGCCAT
 351 GATTAAGTAC CTCACACTGT GGAAGAAAGA GGAGCAGGAG GAGCCCTATG
 401 TCAGCCTCAC CCGAAAGAAG ATGCTAATAG ATGGCGAGGA GGTGCTGATA
 451 GAATGGGAGG TGGGCCACTC CATCCGGACC CTGGCTATGC ACCGCAATGC
 501 CTACAAACGT ATGTCACAGA TCCAAGCCTT CCTGGGCTCA GTGCCCAGC
 551 TGACCTATCA GCTCTATGTG AGCCTGATCT CTGCAGAGGT TCCCCTGGGT
 601 AGAGTTGTGC TAATGGTATT TTCCCTGGTA TCTGTCACCT ATGGGGCCAC
 651 CCTTTGCAAT AGTTGGCTA TCCAGATCAA GTACGATGAC TACAAGATTC
 701 GCCTTGGGCC ACTAGAAGTC CTCTGCATCA CCATCTGGCG GACATTGGAG
 751 ATCACTTCCC GCCTCCTGAT TCTGGTGCTC TTCTCAGCCA CTTTGAAATT
 801 GAAGGCTGTG CCCTTCCTAG TGCTCAACTT CCTGATCATC CTCTTGAGC
 851 CCTGGATTAA GTTCTGAGA AGTGGTGCCC AGATGCCCAA TAACATTGAG
 901 AAAAATTCA GCCGGGTGCG CACTCTGGTG GTCCTGATTT CAGTCACCAT
 951 CCTCTATGCT GGCATCAACT TCTCTTGCTG GTCAGCTTTG CAGTTGAGGT
 1001 TGGCAGACAG AGATCTCGTC GACAAAGGGC AGAACTGGGG ACATATGGGC
 1051 CTGCACTATA GTGTGAGGTT GGTAGAGAAT GTGATCATGG TCTTGTTTTT
 1101 TAAGTTCTTT GGAGTGAAAG TGTTACTGAA TTAAGTGCAT TCCTTGATTG
 1151 CCTTGACGCT CATTATTGCT TATCTGATTT CCATTGACTT CATGCTCCTT
 1201 TTCTTCCAGT ACTTGCATCC ATTGCGCTCA CTCTTACCC ATAATGTAGT
 1251 AGACTACCTC CATTGTGTCT GCTGTCACCA GCACCCTCGG ACCAGGGTTG
 1301 AGAACTCAGA GCCACCCTTT GAGACTGAAG CAAGGCAAAG TGTTGTCTGA

FEATURES:

Start Codon: 1
 Stop Codon: 1348
 3'UTR: 1351

FORM 2 (transcript sequence provided):

1 ATGAACACAA GACCACAACA TTCAGAAAGA ACCTCGACAA TGGACAGAGT
 51 TTATGAAATT CTTGAGGAGC CAAATGTGGA TCCGGTTTCA TCTCTGGAGG
 101 AAGATGTCAT CCGTGGAGCC AACCCCGGAT TACTTTTTCC ATTTAGCATC
 151 CTTTTCTCCA CCTTTTGTGA CTGTGGGGAG GCTGCATCTG CTTTGTACAT
 201 GGTTAGAATC TATCGAAAGA ATAGTGAAAC TTAAGTGAAT ACATACACCT
 251 TTTCTTTCTT TATGTTTTCA TCCATTATGG TCCAGTTGAC CCTCATTTTT
 301 GTCCACAGAG ATCTAGCCAA AGATAAACCG CTATCATTAT TTATGCATCT
 351 AATCCTCTTG GGACCTGTGA TCAGATGTTT GGAGGCCATG ATTAAGTACC
 401 TCACACTGTG GAAGAAAGAG GAGCAGGAGG AGCCCTATGT CAGCCTCACC
 451 CGAAAGAAGA TGCTAATAGA TGGCGAGGAG GTGCTGATAG AATGGGAGGT
 501 GGGCCACTCC ATCCGGACCC TGGCTATGCA CCGCAATGCC TACAAACGTA
 551 TGTACACAGT CCAAGCCTTC CTGGGCTCAG TGCCCCAGCT GACCTATCAG
 601 CTCTATGTGA GCCTGATCTC TGCAGAGGTT CCCCTGGGTA GAGTTGTGCT
 651 AATGGTATTT TCCCTGGTAT CTGTCACCTA TGGGGCCACC CTTTGCAATA
 701 TGTTGGCTAT CCAGATCAAG TACGATGACT ACAAGATTCTG CCTTGGGCCA
 751 CTAGAAGTCC TCTGCATCAC CATCTGGCGG ACATTGGAGA TCACTTCCCG
 801 CCTCCTGATT CTGGTGCTCT TCTCAGCCAC TTTGAAATTG AAGGCTGTGC
 851 CCTTCCTAGT GCTCAACTTC CTGATCATCC TCTTTGAGCC CTGGATTAAAG
 901 TTCTGGAGAA GTGGTGCCCA GATGCCCAAT AACATTGAGA AAAAATTGAG
 951 CCGGGTCGGC ACTCTGGTGG TCCTGATTTT AGTCACCATC CTCTATGCTG
 1001 GCATCAACTT CTCTTGCTGG TCAGCTTTGC AGTTGAGGTT GGCAGACAGA
 1051 GATCTCGTCG ACAAAGGGCA GAACTGGGGA CATATGGGCC TGCATATAG
 1101 TGTGAGGTTG GTAGAGAATG TGATCATGGT CTGGTTTTT AAGTTCTTTG
 1151 GAGTGAAAGT GTTACTGAAT TACTGTCATT CCTTGATTGC CTGTCAGCTC
 1201 ATTATTGCTT ATCTGATTTT CATTTGGCTT ATGCTCCTTT TCTTCCAGTA
 1251 CTTGCATCCA TTGCGCTCAC TCTTCAACCA TAATGTAGTA GACTACCTCC
 1301 ATTGTGTCTG CTGTCACCAG CACCCTCGGA CCAGGGTTGA GAACTCAGAG
 1351 CCACCCTTTG AGACTGAAGC AAGGCAAAGT GTTGTCTGA

HOMOLOGOUS PROTEINS:

Top BLAST Hits:

	Score	E
gi 6502963 gb AAF14527.1 AF155511_1 (AF155511) KX antigen [Mus ...	366	e-100
gi 10835267 ref NP_066569.1 Kell blood group precursor (McLeod...	361	1e-98
gi 2135606 pir I39294 McLeod syndrome-associated protein XK - ...	358	8e-98
gi 3183551 sp P51811 XK_HUMAN MEMBRANE TRANSPORT PROTEIN XK (KX...	358	1e-97
gi 4759330 ref NP_004668.1 Testis-specific XK-related protein ...	76	8e-13

BLAST to dbEST:

	Score	E
gi 1891549 /dataset=dbest /taxon=9606 ...	383	e-104

EXPRESSION INFORMATION FOR MODULATORY USE:

library source:

Expression information from BLAST dbEST hits:

gi|1891549 Germinal center B cells

Expression information from PCR-based tissue screening panels:

Mixed tissue

FORM 1:

```
1 MDRVYEIPEE PNVDPVSSLE EDVIRGANPR FTFPFSILE'S TFLYCGEAAS
51 ALYMVRIYRK NSETYRMTYT FSFFMFSSIM VQLTLIFVHR DLAKDKPLSL
101 FMHLILLGPV IRCLEAMIKY LTLWKKEEQE EPYVSLTRKK MLIDGEEVLI
151 EWEVGHSIRT LAMHRNAYKR MSQIQAF LGS VPQ LTYQLYV SLISAEVPLG
201 RVVLMVFSLV SVTYGATLCN MLAIQIKYDD YKIRLGPLEV LCITIWR TLE
251 ITSRL LILVL FSATLKLKAV PFLVLN FLII LFEPWIKFWR SGAQMPNNIE
301 KNFSRVGTLV VLISVTILYA GINFSCWSAL QRLAD RDLV DKGQNWGHMG
351 LHYSVRLVEN VIMVLVFKFF GVKVLLNYCH SLIALQLIIA YLISIDFMLL
401 FFQYLHPLRS LFTHNVDYL HVCVCCHQHPR TRVENSEPPF ETEARQSVV
```

FORM 2:

```
1 MNTRPQHSE R TSTMDRVYEI PEEPNVDPVS SLEEDVIRGA NPRFTFPFSI
51 LFSTFLYCGE AASALYMVRI YRKNSETYWM TYTFSFFMFS SIMVQLTLIF
101 VHRDLAKDKP LSLFMHLILL GPVIRCLEAM IKYLT LWKKE EQE EPYVSLT
151 RKKMLIDGEE V LIEWEVGHS IRTLAMHRNA YKRMSQIQAF LGSVPQ LTYQ
201 LYVSLISAEV PLGRVVL MVF SLVSVTYGAT LCNMLAIQIK YDDYKIRLGP
251 LEVLCIT IWR TLEITSRL LI LVLFSATLKL KAVPFLVLNF LIILFEPWIK
301 FWRSGAQMPN NIEKNFSRVG TLVVLISVTI LYAGINFSCW SALQRLADR
351 DLVDKGQNWG HMGLHYSVRL VENVIMVLVF KFFGVKVLLN YCHSLIALQL
401 IIAYLISIGF MLLFFQYLHP LRS LFTHNVV DYLHCVCCHQ HPRTRVENSE
451 PPFETEARQS VV
```

FEATURES:

Functional domains and key regions:

[1] PDOC00001 PS00001 ASN_GLYCOSYLATION
N-glycosylation site

Number of matches: 2

```
1 302-305 NFSR
2 323-326 NFSC
```

[2] PDOC00004 PS00004 CAMP_PHOSPHO_SITE
cAMP- and cGMP-dependent protein kinase phosphorylation site

Number of matches: 2

```
1 59-62 RKNS
2 169-172 KRMS
```

[3] PDOC00005 PS00005 PKC_PHOSPHO_SITE
Protein kinase C phosphorylation site

Number of matches: 6

```
1 64-66 TYR
2 137-139 TRK
3 157-159 SIR
4 252-254 TSR
5 264-266 TLK
6 354-356 SVR
```

[4] PDOC00006 PS00006 CK2_PHOSPHO_SITE
Casein kinase II phosphorylation site

Number of matches: 3

```
1 17-20 SSLE
2 18-21 SLEE
3 431-434 TRVE
```

[5] PDOC00007 PS00007 TYR_PHOSPHO_SITE
Tyrosine kinase phosphorylation site

126-133 KEEQEEPY

[6] PDOC00008 PS00008 MYRISTYL
N-myristoylation site

Number of matches: 4

1	215-220	GATLCN
2	321-326	GINFSC
3	343-348	GQNWGH
4	350-355	GLHYSV

[7] PDOC00029 PS00029 LEUCINE_ZIPPER
Leucine zipper pattern

100-121 LFMHLILLGPVIRCLEAMIKYL

Membrane spanning structure and domains:

Helix	Begin	End	Score	Certainty
1	36	56	1.443	Certain
2	74	94	2.084	Certain
3	102	122	0.920	Putative
4	181	201	0.811	Putative
5	208	228	1.744	Certain
6	273	293	1.234	Certain
7	312	332	1.785	Certain
8	366	386	0.828	Putative
9	389	409	1.497	Certain

BLAST Alignment to Top Hit:

>gi|6502963|gb|AAF14527.1|AF155511_1 (AF155511) KX antigen [Mus musculus] Length = 446
Score = 366 bits (930), Expect = e-106
Identities = 179/411 (43%), Positives = 265/411 (63%), Gaps = 11/411 (2%)

Query: 33 FPFSLFSTFLYCGEAAALYMVRIYRKSETYRMTYTFSSFFMFSSIMVQLTLFVHRDL 92
FP S++ S FL+ E A+ALY+ YR + T F + +VQ TL+FVHRDL
Sbjct: 3 FPASVIASVFLFAETAALYLSSTYRSAGDRMWQVLTLLFSLMPCALVQFTLLFVHRDL 62

Query: 93 AKDKPLSLFMHLILLGPVIRCLEAMIKYLTWKKKEEQEPPYVSLTRKKMLI-DGEEVLIE 151
++D+PL+L MHL+ LGP+ RC E Y + ++ EEPYVS+T+K+ + DG +E
Sbjct: 63 SRDRPLALLMHLLQLGPLYRCCEVFCEIYC---QSDQNEEPYVSITKKRQMPKDGGLSEEVE 119

Query: 152 WEVGHSIRTLAMHRNAYKRMSQIQAFGLGSVPQLTYQLYVSLISAEVPLGRVVLVMVFSLSV 211
EVG + L HR+A+ R S IQAFGLGS PQLT QLY++++ + GR +M SL+S
Sbjct: 120 KEVGQAEGKLITHRSASFRRASVIQAFGLGSAPQLTLQLYITVLEQNITTGRCFIMTSLLS 179

Query: 212 VTYGATLCNMLAIQIKYDDYKIRLGPLEVLCITIWRTLEITSRLILVLFSATLKLKAVP 271
+ YGA CN+LAI+IKYD+Y++++ PL +CI +WR+ EI +R+++LVLF++ LK+ V
Sbjct: 180 IVYGALRCNILAIKIKYDEYEVKVKPLAYVCIFLWRSFEIATRIVIVLVLF+SVLKIWVVA 239

Query: 272 FLVLNFLIILFEPWIKFWRSGAQMPNNIEKNFSRVGTLVVLISVTILYAGINFSCWSALQ 331
+++NF PWI FW SG+ P NIEK SRVGT +VL +T+LYAGIN CWSA+Q
Sbjct: 240 VILVNFFSFFLYPWIVFWCSGSPFENIEKALSRVGT+IVLCFLTLTYAGINMFCWSAVQ 299

Query: 332 LRLADRDLDVKGQNWGHMGLHYSVRLVENVMVLVFKFFGVKVLNLYCHSLIALQLIIAY 391
L++ + +L+ K QNW + ++Y R +EN +++L++ FF + + C L+ LQL+I Y
Sbjct: 300 LKIDNPELISKSNWYRLLIYYMTRFIENSVLLLLWYFFKTDIYMYVCAPLLILQLLIGY 359

Query: 392 LISIDFMLLFFQYLHPLRSLFTHNVVD---YLHCVCCHQHPRTRVENSEP 438
I FML+F+Q+ HP + LF+ +V + L C C R ++SEP
Sbjct: 360 CTGILFMLVFYQFFHPCKKLFSSSVSESFRALLRCACWSS---LRRKSSEP 407

ALIGNMENT OF FORM 1 AND FORM 2:

>FORM 2

Length = 462 (Length of FORM 1 = 449)
Score = 900 bits (2301), Expect = 0.0
Identities = 447/449 (99%), Positives = 447/449 (99%)

FORM 1: 1 MDRVYEIPEEPNVDVPVSSLEEDVIRGANPRFTFPFSILFSTFLYCGEAAALYMVRIYRK 60
MDRVYEIPEEPNVDVPVSSLEEDVIRGANPRFTFPFSILFSTFLYCGEAAALYMVRIYRK
FORM 2: 14 MDRVYEIPEEPNVDVPVSSLEEDVIRGANPRFTFPFSILFSTFLYCGEAAALYMVRIYRK 73

FORM 1: 61 NSETYRMTYTFSSFFMFSSIMVQLTLFVHRDLAKDKPLSLFMHLILLGPVIRCLEAMIKY 120
NSETY MTYTFSSFFMFSSIMVQLTLFVHRDLAKDKPLSLFMHLILLGPVIRCLEAMIKY
FORM 2: 74 NSETYWMTYTFSSFFMFSSIMVQLTLFVHRDLAKDKPLSLFMHLILLGPVIRCLEAMIKY 133

FORM 1: 121 LTLWKKEEQEPPYVSLTRKKMLIDGEEVLIEWEVGHSIRTLAMHRNAYKRMSQIQAFGLGS 180
LTLWKKEEQEPPYVSLTRKKMLIDGEEVLIEWEVGHSIRTLAMHRNAYKRMSQIQAFGLGS
FORM 2: 134 LTLWKKEEQEPPYVSLTRKKMLIDGEEVLIEWEVGHSIRTLAMHRNAYKRMSQIQAFGLGS 193

FORM 1: 181 VPQLTYQLYVSLISAEVPLGRVVLVMVFSLSVSVTYGATLCNMLAIQIKYDDYKIRLGPLEV 240
VPQLTYQLYVSLISAEVPLGRVVLVMVFSLSVSVTYGATLCNMLAIQIKYDDYKIRLGPLEV
FORM 2: 194 VPQLTYQLYVSLISAEVPLGRVVLVMVFSLSVSVTYGATLCNMLAIQIKYDDYKIRLGPLEV 253

FORM 1: 241 LCITIWRTLEITSRLILVLFSATLKLKAVPFLVLNFLIILFEPWIKFWRSGAQMPNNIE 300
LCITIWRTLEITSRLILVLFSATLKLKAVPFLVLNFLIILFEPWIKFWRSGAQMPNNIE
FORM 2: 254 LCITIWRTLEITSRLILVLFSATLKLKAVPFLVLNFLIILFEPWIKFWRSGAQMPNNIE 313

FORM 1: 301 KNFSRVGTLVVLISVTILYAGINFSCWSALQLRLADRDLDVKGQNWGHMGLHYSVRLVEN 360
KNFSRVGTLVVLISVTILYAGINFSCWSALQLRLADRDLDVKGQNWGHMGLHYSVRLVEN

FORM 2: 314 KNFSRVGTLVVLISVTILYAGINFSCWSALQLRLADRDLDKQNWGHMGLHYSVRLVEN 373

FORM 1: 361 VIMVLVFKFFGVKVLNLYCHSLIALQLIIAYLISIDFMLLFFQYLHPLRSLFTHNVVDYL 420
VIMVLVFKFFGVKVLNLYCHSLIALQLIIAYLISI FMLLFFQYLHPLRSLFTHNVVDYL

FORM 2: 374 VIMVLVFKFFGVKVLNLYCHSLIALQLIIAYLISIGFMLLFFQYLHPLRSLFTHNVVDYL 433

FORM 1: 421 HCVCCCHQHPRTRVENSEPPFETEARQSVV 449
HCVCCCHQHPRTRVENSEPPFETEARQSVV

FORM 2: 434 HCVCCCHQHPRTRVENSEPPFETEARQSVV 462

Hmmer search results (Pfam):

Model	Description	Score	E-value	N
CE00306	E00306 Membrane_transport_protein_XK	390.8	1.3e-113	1

Parsed for domains:

Model	Domain	seq-f	seq-t	hmm-f	hmm-t	score	E-value
CE00306	1/1	31	416	..	1 384 [.	390.8	1.3e-113

```

1 TATTATTATT ATTATTAAGA CGTAATCTTG CTCTGTTGCC CAGGCTGGAG
51 TGCAGTGGCG TGATCTCAGC TCACTGCAAC CTCTGCCGTC CGGGTTCAAG
101 TTTTCTCTCT GCCTCAGCCT CCTGAGTAGC TGGGATTACA GTCACGCACC
151 ACCACGACCA GCTGATTTTT GTATTTTTAG TAGAGATGGG GTTTCACCAC
201 GTTGGCCAGG CTGGTTTCGA ACTCCTGACC TCAAGTGATC TGCCTGCCTC
251 AGCCTCCCAA AGTGCTGGGA TTACAGGCGT GAACCACTGT GCCTGGCCTT
301 CATCTATATT ATTACCAGGA GGCAGATGTG TTCTCTTTTT CTCTGAGGTT
351 TAGAATTATG CAAATGAAGA TATGAAAACA AAAGCTCAGT GAGGTGGGGA
401 GGATTACACT TAAGAATACA GGTAATTTTC AAAGCTCTTT AAGACACCCC
451 TCTCAGTTTT TACTAACAGC TCTCTCTTGG CTCTTTGCCA GTCTGTTTAG
501 AATTGTGGCAG CTCTTCATAA CCTTTCAACC AAAGACCTGT AAGTTCATTC
551 TAAAGCTCCT ATCCTGGCCT CATTTTGCAA GTGGAGAAAT CAAGGCATAA
601 AATATGAGCT TTCAGTGTCT GTGGGCTGAC CTTGAGTCTT GACCTTTATC
651 CTGTTCTATC TTCCCTCCGC CGAAAACCTC GACCCTATTC CTCCCAGGTT
701 CCCCTTTCAT GATATTATCT GGAGGGCAAT AGGACCTAGG GAGGTCCAC
751 CCTGCGGCGG AGGGAGACAC ACCTGCCTAA CAGCGTGGGT AGAGTGAGTG
801 TTGAAGCAAG TCACTTAACT AGTTAGGGAG GCGGGGGTAG AAGTGGGGGC
851 CTGCTGCTCC TAGGGAGGAG TAAAGCTGTG GCTCCTGCCT GGGTCTGGAG
901 GTGGTGGTCA GAAGTGCTTC TGAAGAGCGG CCAAAGCCCC TTTTGTCCC
951 GCCACTCCAC AACGAGCATC CCTCGGCTGG CCGCCTGCCC GGGAACTCTC
1001 CGGCTGGTTT TGTTTGGCCG CAGCCGTCCC GCCCATCTCG CCCGCCCCCG
1051 CCGTCCCGGT GCCTTAGTTC TTGAAGCTGC CGACCTCTCG CAGCTGGAAT
1101 CGCAGACCAG GCAGGACCCT GGCAGCAGAC GCGCTCCAAG AGTTTGCGCA
1151 CCTCCGTCCA GCCAGTTTGG CGCCCCGCAC ATCGTGCTC TACTAGCAA
1201 AGTTTCTCCG AGGAGAAGCA GCCCTCCAG CCTTTTCTTC ATCCTGTAGA
1251 GCGAGCGCGC TCTGCTTCTG TCCCTCAACA CTGCATTCTG AGACAGGGTG
1301 GTGACAATAC TCCACTCCCG GGCCAGGCGG TCTTGGGGGC GGGGCTTGGG
1351 GGAATCCGAG GAGCTATCCT GAGAACCCTG GACTCGGCAA AGGTCTGAG
1401 AGCGCGCAGG TGAGCGGGCC AGCTGATAGC TACAGCCTAG CAATAGCTAG
1451 GATACCTAGG CACTGAACTG AATCCCTCT TCTGCCCTCC TTCTTCTGCG
1501 CCCGCTCTTC TGCCCTGGCT CAGCTCTCCG CTGACTTGAG AGGACACACT
1551 GGTGAGGACT CTTTGTGAGG AGCTGCTGAG TGTGCGTGCC CCCGACAGAT
1601 CGGCTACACC CTGCCTGAGG GGCTGCGAAA GGAGCCGCCA CGGAAGCCGC
1651 TGTTCTCATG ACTCTTCACG TCCCTGGAGT TGGACTCTGG ATGGGGCGCT
1701 GGGATGCTTG CTTTTGTCTT GTTCAAGTTT CACAGCAAGT ATGTTGACGA
1751 TTGGAATCGG GGCCAATCAA GAGTCAAGTT CAAAGTGGTA CTCCTGGGCT
1801 TTCCATCCCA GACTCCAAGT CGAATCTGAG TCTAGAAGAG AGCGGTTTCT
1851 TGCTCTAACT AGTGAATCTC TGTTCCCAA CTGGACTTGA CAGAGCTCTC
1901 CTCACCTATA CTTGGACTGT AGCGGCCATA GGGTTCTCTT GGGGATGGGT
1951 GGGAGGGTGC TATGAACACA AGACCACAAC ATTCAAGAAAG AACCTCGACA
2001 ATGGACAGAG TTTATGAAAT TCCTGAGGAG CCAATGTGG ATCCGGTTTC
2051 ATCTCTGGAG GAAGATGTCA TCCGTGGAGC CAACCCCGA TTTACTTTTC
2101 CATTTAGCAT CTTTTCTCC ACCTTTTGT ACTGTGGGGA GGCTGCATCT
2151 GCTTTGTACA TGGTTAGAAT CTATCGAAAG AATAGTGAAA CTTACTGGAT
2201 GACATACACC TTTTCTTCT TTATGTTTT ATCCATTATG GTCCAGTTGA
2251 CCCTCATTTT TGTCCACAGA GATCTAGCCA AAGATAAACC GCTATCATT
2301 TTTATGCATC TAATCCTCTT GGGACCTGTT ATCAGGTGAG CAACTTTTAA
2351 ATCTTTTCTT TACCCCTTA ACCCCACCCC AGACTTGGG AGAGAAAGAT
2401 GAAAGATTTA CAAGATGGAT ACTATGGCTC TAATCAATTC TCTCATTTCC
2451 TCCACTCTC GGCTTCCCTG TCTACCATT AGAAACTTA CCTGAAATCT
2501 TAAATGCCAC CATGATGAAC ATGTGGTATG TACTTGTGTT CAAAACAAT
2551 GAACGATGCT ATTTGGGCTG TGTAAGTAG AATGGGAACA ACAAGACGTG
2601 ATCACCCTGT GCATGAAGGC CATAGCTGCA GAGTGTGTAA TTTTATTTAA
2651 AAAAATTTTT TTTTCTGAGA CAAGGTCTTG CTCTGCCTCC CAGGCTACAG
2701 TGCAGTGGTG CGATCATGGC TCACTGCAGC CTTGATCTCC TGGGATCAAG
2751 CGAACCTCCC ACCTCAGCCT CCAAGTAGCT GGGACCAAAG GAATGTGTCA
2801 CCATGCCCTG TTAATTAATA AAAAATTTTT ATAGCCGGT TGTGGTGGCT
2851 CATGCCTGTA ATCCAGCAC TTTGGGAGG TGAGGCGGGT GGATCACCTG
2901 AGGTGAGGAG TTCAAGACCA GCTGGCCAAC ATGGTGAAAC CCCTGTCTCT
2951 ACTAAAAATC AGCTGGGTGT GGTGGCGCAT ATCTGTAATC CCAGCTACTC
3001 TGGTGGCTGA GGCAGGAGAA TCACTTGAAC CCGGAAGGTA GAGGTGACAG
3051 TGAGCCAAGA TCGGTGCCAC TGCACTCCAG CCTGGGCGAT AGAGTGAGAC
3101 TCCATCTCAA AAAAAAAAAA ATTTTTTTTG TAGAGACGGG ATCTCGTTAT

```

FIGURE 3, page 1 of 8

3151	GTAGACTGGG	CTCAAGTGAT	CTTCCTGCCT	CAGCCTCCCA	AAGTGAGCCA
3201	CCACGCCTGG	TCTGAGTGTG	TAATTTTGAC	TCTACCTTTT	TGGATGCTTT
3251	GTAAATTGGA	TAAAAGTTTC	TTTACCCTGA	GCTGCTGGG	CTGGTGCTAC
3301	TGCCATTTTC	AAATTTTCCA	GAGTAATGTG	ACATCTGGAA	ACTATTTTAA
3351	ACCATCTGTG	GTAATCTGTA	CCCCAACCCA	ATATAGTTCA	GTCTCTGTCT
3401	GGTTTATCAG	TTTCTTATTT	ATCTCTTTGT	ATATTCTCTG	AATAAAGATA
3451	CGAAGTTGGG	AGGGGGCAAA	GGAAGGCAGT	TCATCTCTCT	ATGTGGATGC
3501	AGTAGCACAA	TTTAATAGTA	TCAAGTATTT	CCATTCCAGAT	TGCCTTGAAG
3551	TGGAAGAAGT	GCACTTAATC	CTAGCGAGAT	AGGCACCTGT	GTCACAGTC
3601	TCATCTGGAT	GCTATGGGGT	TTTCAAGGTA	GAGAGATGTT	GCAAAACTTA
3651	TGAGTCTCAG	AGTAAGGAAT	GGACCAAGTT	TGTCCTGATT	GCGAGAGAGG
3701	CAGACAACTG	CAGTCAAGCCG	AGGAATATGG	GTCAGAGTGT	TGCAATGGGA
3751	AGATACCTCA	TCATTAGACA	ACTAAAAAGT	CTGTGAAACT	AATTAAGGAT
3801	GGAACCTACT	CCTTTATAAA	ATTTCATATC	TGTACACATG	TATAATTTTT
3851	ATTTGTCACT	TATACCTCAA	TAAGGCCAAA	AAAATTTTTT	ATCAATAAAT
3901	TTTTAAGTGG	GGAGGAATCG	ATTAGGCTCT	ATCAGAGAGA	ATATGGGATA
3951	TCAATGGAAA	CAGTGGCCTG	AAATTTGGAG	TCTAGTCTTC	CGCCTGTCTAT
4001	TGACTGGTTG	TGTTGTCCTG	GTAATACTCT	TGAAGATGGC	TTACACGGAA
4051	GGCATATAGA	GTTCCCTCAT	CTGTAAAGCA	AATGGGTTAG	TCTAAATCAT
4101	GGGTCTCAA	CTCAAACACT	TGCAGGGACC	AGGCAGGTAT	CATAAATGAA
4151	TGAAGCAGGC	CTAGTATAAG	AAAAACAGT	AGCCTTGTGT	GAGATGATAA
4201	ATGGAAACAA	AGTCTCAGAG	AAATACTGAG	GAGTAGTGAG	TACCATGGTA
4251	ATCTGAAATC	TTCATGACCT	GCCTGAAGGA	GGTAGCCCCT	CTAGAGCCCT
4301	GGCGCATTGT	TTCCATGTTG	GAATTCAGAC	CCGATATTGC	CAGATGCCAT
4351	AACTTTTTCG	GAGATGCTCC	CAAGACAGGA	TTTTTATATG	AAATGCTCATG
4401	ATTTTAAATT	TTACACAGCTG	ACTAAAAACA	TAACAACAAC	AACACAGGAT
4451	GGACCAAACC	ATATCTGTTG	GTCAGATATA	ACTCAGCTGG	CCTATATGCA
4501	TCTTTGGACT	GGGTGATGTA	AAGTCTCTTT	ACGGTTCTAA	ATCTTTGAAG
4551	TTAAGCTGTA	AAAGGAAGAC	CTCATCTTTGA	CCTTGAAACC	AAGAAATTTA
4601	AAGTTGTGAC	TACAGGAGCA	AATAAACCAT	TCATCCCTCC	TTTTTCAAAT
4651	ACAAATATATT	GAGTTAACCA	ATCGAAAACT	CTCAGAGATC	AAATTTTCAGA
4701	AAGTACCCAG	CTGCACCCTC	CCCTCTTTTT	GACTTCCTTT	GTTTGCTTTG
4751	TGAACCCTCT	GTGTAGAGTG	TTGAGTACTG	TTTTTCATTT	TTGTTGTTTA
4801	GCTTCCACTA	GAAATGATTG	GGAAGCATTT	ATAACCTCAG	GCAGCTTAGC
4851	CCACAGCAGA	GAAAAGATAA	AAACTCATAA	ATTATACTCT	GGATTGCGTT
4901	ATTTTCAAGG	CCAATTACTT	GTTAGATAGG	TAGGAACCTG	ATTAGTGTTA
4951	TCAGGCACAT	GAAAGTGCTT	GTAGAGTCTG	GGTGCCCTAG	ATGAATGCA
5001	AGCATACTTC	CGAAATGAAA	ATGTACTCTA	ATTTATTGAA	GCTTATAAAT
5051	GGACAAACAC	CCTTACTTAA	ACCAGAAAAT	AGCCCTGAGA	ATAGAAACAG
5101	AACATTTATG	TAAATGTAAA	CGGAACATTT	CATGCCACCA	CCTTCTCCAA
5151	TACTGTTCTC	CAATTTAGCA	ATAGTACTGA	TGGGTGGGG	TTAAAACTA
5201	AAATTTTTTC	TTGAAAAATC	ACTTATGCAG	AACAAGAATA	GGAAAAAGT
5251	TTTGCTTTTT	TCTCTCTGTT	CTTTCTTTTG	ATCTTTTTCT	TTCCCAGGTC
5301	TGTAGTTTTG	TCCCTAGAAG	GTGACAAATT	CAAACTACAT	GCTTCAGAGT
5351	GGTACACATG	CATCAGTCTT	AGGGTGATCT	ATGGAGACTG	GCAGCCAGCA
5401	TATGTTCCAA	ATTTTCCTAT	CAGGAACTAA	AGGCTAGAGA	GCATATCAAC
5451	CTCTGGGCTT	GTCTTTGGTC	TACTTTTCTG	TTAAATTTCA	TTGCTGTTAT
5501	TATTATCCTC	TCCTCCCATA	ATTGCTTACC	CTGTATTATT	TTCTTCCTTC
5551	TTATTCTTTT	ATTTACTCAG	CAAAATTTTC	TCAAATACCT	ACTAAGTGAT
5601	AAGAGCTGTA	AACAGATATA	ATACAAACCT	TGACCTCAGT	CTCTTGGGCA
5651	AGACGTGTTA	ATGTCCACTA	CAAAATGTTCT	TACTAGTCAT	AAGTAGTCCA
5701	CAGTTTTTAT	TCATTAAAGG	TGAGTGGCGA	AGTGGTAACT	CAGGTGTTCC
5751	AGTAACAAGA	ATGTTCTAGT	TGCTTCTCTT	CCACTTACCA	CATCAGAACT
5801	GCTAAAGACT	TCTGATTTTG	ATGGGGGAGG	TGGGAGGGGC	AGAGCAGGAA
5851	ATGTCATCTT	ACCTTTATTC	CAGGATGATG	AGGCTTTTCA	AAGGATGTTT
5901	TTCTCTTCGT	AAAGAAAGAA	TCCAGTTTAA	AAGGCTTTTT	TCCACAAACA
5951	GGACAAAGAG	CACAAAAAGT	AACTATTACA	GTGATCTTTC	GAGGGCCTAG
6001	TTATGTAGTT	CATTCAGGTT	TGAGTTGTCT	TCTTTTAAAGT	ACTTTTGTTG
6051	CTTTGATGGC	TTCTGTGTGA	TATGAGATAT	TTTTTTTCTT	CTGATCTGTC
6101	CCAAGACTTT	TTGGCTGAGA	TATGGTTGTG	AGCCCTTTCT	TGAAAAAGCA
6151	GAATCTGGCC	AGGCGCAGTG	GCTCATGCCT	GTAATCTCAG	CACTTTGGGA
6201	AGCTGAGGTG	GGTGGATCAC	CTGAGGCTCAG	GAGTTCAAGA	CCAGCCTGGC
6251	CAACATGGTG	AAAACCCGTC	TCTACTAAAA	ATACAAAAAA	AAAAAAACAC


```

6301 TTAGCCGGAC ATGGTGGCAC ATGCCTGTAA TCCCAGCTAC TCAGGAGGCT
6351 GAGGCAGGAG AATCGCTTGA ACCCAGGAGG CAGAGGTTAC AGTGAGCTGA
6401 GATCGCGCCA GTGCACTCCA GCCTGGGCGA CAGAGCAAGA CTCTGTCTCA
6451 AAAAAAAAAA AAAAAAGAAA GAAAGAAAAA GAAAAAGCAG AATCTAAAC
6501 TTTGGTTATG GAGCTGAATG CTTTGAGGGA GGAATGCTTT ACCTCACGAA
6551 TTTGAGGTAA GAAACAGGG CTTTGGAAC CTTCAATTAT TTGCTAGGAA
6601 AACAGTATCG ACTTAATACC TTTGTGTTCA AGGCACTTTT CTACCTGCCA
6651 CAGGCCTATT CTTAAAAAGA CAAAACAATT CCTCGAGTCC TCAAACAAGT
6701 ACTTCTGAAA CAGTGTCTT AGGTCAGTCG ATGACTGAAC AAAAATGGAT
6751 TTAGATTTCAT GTAACCTGTA GAAGGCATGA TCCACCCTTT GACTTATGAG
6801 AAATGATCAG AACAGAAGAG AGAAAAAGAC AAAAAGTAGT GCAGGCTGGC
6851 CATGGTGTCT CACACGTGTG ATCCCAGCAC TTTAGGATCC CAGCACTTTG
6901 GGTCAAGTAA GTAGGATTGC TTGAGCCAG GAGTTTGAGA CCAGTCTGGG
6951 CAACATGTCT AGATCTCCTC TCTACACAAA TTAAAAATAG CTGGCATGGT
7001 GGCATGCGCC TGTAGTCCTA GCTACTCAGA AGGCTGAGGT GGGAGGATCA
7051 TTTGAGCCTA GGAGGTCAA GCTGCAATGA ATTATGATTG TGCCACTGCA
7101 CTCCAGCCAG GGTGATGGAG TAAGACCTTG TCTCAAAAT AAAATAAAGT
7151 AGCACAACCT CCCCAAGTTA TTTTTTTCCC TACTACAAC CTCCCTTCCC
7201 AGGACAGCTT AGTTAAGTT GCATGATGCT TTACTTCTGC AGATGTTTGG
7251 AGGCCATGAT TAAGTACCTC ACACTGTGGA AGAAAGAGGA GCAGGAGGAG
7301 CCCTATGTCA GCCTCACCCG AAAGAAGATG CTAATAGATG GCGAGGAGGT
7351 GCTGATAGAA TGGGAGGTGG GCCACTCCAT CCGGACCTG GCTATGCACC
7401 GCAATGCCTA CAAACGTATG TCACAGATCC AAGCCTTCCT GGGCTCAGTG
7451 CCCAGCTGA CCTATCAGCT CTATGTGAGC CTGATCTCTG CAGAGGTTCC
7501 CCTGGGTAGA GGTGAGTGGG GTCAGGAGAG GGGAGGGCTC CAGTTAAATC
7551 AAGGGTCTTA GAAGTCTAGA CCCAAGCTGT CTAATAAACT GGCCACTAGC
7601 TTCATGTGGC TATTTAACTT AAAATTAAAT AAAATTAAAA ACTTGTTTCA
7651 TAATACTAGC TACATTTCAA GTTCTCAGCA GCCGTGTGTT GCTAGCAACT
7701 ACTGTATTGG ATGGCACAGG TATAAACATT TCCATCATCA CAGAAAGTTC
7751 TATCGGACAG CACTGGGAGA TAGTTAAATA ACTTGTGGAG TCAGACATCT
7801 CAAGCCTGCC AGATTTCTTA AACAGGTAAG CTGTTTAGAC TAAAAATGTC
7851 ACAGATAAAC CTTCTTGGG CCCAGAAGAA GCTAGTAATA CCAGCACTCA
7901 GTAGGATATT TTCCCTTGCC CAAAATGTTT AAATTATGCT GTTGTTTGT
7951 TTGTTTAAGG ATGGCAGTCT TTAATAAGAG GTTCCCAAAT AGTACTGATC
8001 ATCAGAATCA TGTGATGAGC TTCTTTTTGA AATTATATTC ACTCCCAGA
8051 CTTGAATCAA TCTTAATATG TATTTCTAAA AGGTACCCAG TTGATTTTGA
8101 TCAGCCACAT TTGGGAACCA ATGATTTAAT CATTTCTGCT AATGCCAGTG
8151 GAGAGAAAGA AAAGGAGCGT GGGCTGGGCA CGGTGGTTCA AGCCTGTAAT
8201 CCCAGCACTT TGGGAGGCCA AGGCGGGTGG ATCACAAGGT CAGGAGATTG
8251 AGACCATCCT GGCTAACATG ATGAAAACCC GTCTGTACTA AAAATACAAA
8301 AAATTAGCCG GCGTGGTGG CAGGTGCCTG TAGTCCTAGC TACTCGGGAG
8351 GCTGAGGCAG GAGAATGGCG TGAACCTGGG AGGCGGAGCT TGCAGTGAGC
8401 CGAGATCGCG TCACTGCACT CCAGCCTGGG TGACAGAGCA AGACTCCGTC
8451 TCAAAAAAAA AAAAAAAA AAAAAAAAAA GCGTGGGGTT AATACTAATG
8501 AGAGTCAGGC CTGGACCAAG TTCTGACCTT CACTGTGATC TTTGGAGGAA
8551 GTTACAAAGC AATCACTGAC CTAATTTCCC ACTTGTAGAA GAGGGATCCT
8601 GAAATGAGTA AGACCTCTAG CAGAAGATGA AATGTGAGTC AGTGTTTTCA
8651 AAGTTGAGAT AAATTGTTGT TAATGAATTT TAACAGCCTG AGATTTGCTT
8701 CATCTGCTTG GGCAGGCACT GGTATAGGTG TGGGTACAGG TTTGGACCAT
8751 TTCCTATTAG ATTCTAACCC TGTTTGCCAA AGTCCCATGT CTCAAATAAG
8801 GTAAGGAGAA AATTGGCCCT CTTTGTCTT TTTTCCCCAC TCAGAAATTGT
8851 TCTTGAAGTT CTGTTGGTCT TGAAGCTTT CACATACATA GTAGTTTGAG
8901 GAGAAACTC TTTGGAATG ATGATGCTT TCCTTTAAAT CATCTAATAA
8951 AAATAGGTGT ACATTACGGC TGGGCATGAT GGCTCACGCC TGTAATCCTA
9001 ACACCTTGGG AGGCCAAGAC AGGCAGATCA CTTGAAGTTG GGAGTTCAAA
9051 ACCAGCCTGG CCAACATGTT GAAACCCCAT CTCTACTAAA AACACAAAAA
9101 AAATCAAGGA TGGGCATGGT GGCTGATACC TGTAATCCCA GCACTTTGGG
9151 AAGCCGAGGC AGGTGGATCA CCTGAGGTCA GGAGGTTGAG ACCAGCCTGG
9201 CCTGGCGAAA CTCTGTCTCT ACTAAAAATA CAAAAATTAG CCGGGTATGG
9251 TGGTGGATGC TTATAATCCC AGCTACTTGG GAGGCTGATG CATGAGAATC
9301 ACTTGAACCT GGGAGCCGAG ATCTCACCGT TGCACCTCAG CCTGGGCAAC
9351 AGAGCGAGAC TCTGTCTCAA AAAAAAATTC AGCCAGGCGT GGTGGTGGGT
9401 GCCTGTAATC CCAGCTACTT GGGAGGCTGA AGCAGGAGAA TTGCTTGAAC

```

FIGURE 3, page 3 of 8

```

9451 CTGGGAGGTG GAGGTTGCAG TGAGCTGAGA CTGCACCACT GCACCCAGC
9501 CTGGGCGACA GAGGGAGACT CCCGTCTCAT AAATAAATAA ATAAATAACA
9551 AAAGTAATAC ATGCACAAAA TGACATATAA GTAATTGTAT TTGCACAGAA
9601 AATTTCTGGA AACTATGCAA GAAACTACCT CTGCGGAGTG GGAATGAAAA
9651 GTCAGCAGTC TTACTTTTAA AAATTCCTCT GTATGGTTTG AAAATTTTTT
9701 TTGTGATCAT GCATTACTAG TTTTGGTCTT TATCTTTTTT TAATTACAAA
9751 AGTCAGACAT GGTTATAGTA AAAATTAAAA ACCATACAGA ATAGATATAA
9801 AATAGGAAAC GTAATCTCAC TCCCCAAAGA TAACCTCTGT TAATCATCCA
9851 GTATATATCC TTCTGGACTT ATTTTTACTA TGTAACATA AACATACATA
9901 CAATATATAT TGTACATGTT TTTGCCCAA AATGGACTGT ATGAAACATT
9951 CTGTCAACAA AGTATTTTTC AAAAGTACAG TATGCCAGTA TGTCTTTTCT
10001 CAAGTTATTT ATATATACAT GTATAACAAT AATAAATATA TAATATACAT
10051 TTCCTTTATA TGAATTAGAC TATTTTATTT CTCCTAATTT TCTATTGATA
10101 GGATTCTATT GATTGTCTCA AAAAGGAAAA AAAAAGGTAG CACAACCTCC
10151 CCTAGTTATT TTTTCCCTC ATTACAACCT CCCTTCTCAG GACAGCTTTA
10201 GTTAAGTTCC CATGATGCTT TACTTCTGCA GATGTTTGA GGCCATGAGT
10251 AAGGACTTCA CACTGTGGAA GAAAGAGGGG CAGCAGGAGC CCTATGTCAG
10301 CCTCACAAAT TTTAATTTTT CACAAAAAAG TTGTTTCTTA ATTGCAAATT
10351 ATGCCACAGT AAACATCTTT ATAAATACCT GTGTACATGA ATGAGACTTT
10401 GTAGGATAAA TTTATAGCAG TAGAATTGCT GGGTGGGAAG ATATGTATGT
10451 TTAATAATTT ATTGATATTG CCAAATAACT CTTCCAAAAA GATATATGAA
10501 TTTATACTCT CACCAACAGT ATACAAATGT GCCTGTTTCT GTTTCTTCAT
10551 ATCTTAAACT CAATATCTTT TATTTGTATA ATTATAAAAT AATTGGCTTT
10601 TAAAATAATT GACTTTTAAA ATAATTCGCT TTCTTTGGTT ATGAATGAAG
10651 CTGAGCATCT TTTTGTGTTT GGTCAATGTG TGTCTGTGA ATTGCTTGT
10701 TATATATTTT ACTCGCTTTT TCTAGTGGGT TGTCTTTTTC ATATTAATTT
10751 TTAGGAGCTA TTTACTTATT CTTGTTATTA ATCCTTTCTC TGCTGTGAAT
10801 ATGTATGCAT ATATTTGTAT AATTTTTTGC TTGTACATAC ACACATTTTA
10851 AATATGTATA TACATGTAT ACGTGTATA TGTGTGTGAT ATATTTAATA
10901 TCCACAATAC ACTTTGTAGT ATCTTCTGGC ATTCTGAAGT ATTACATTTT
10951 TATGTATTCA AATTCCTTAT TGCTTTTGA TTTTGTGCCT TTCTTACAAA
11001 GGCCTATCAC ATCTCTCATC TGGTAGAACA ATTTTCCCCA ATCTTTTAAG
11051 TAGATTAATT TTCAAGATAA TTTTTTAATT CATCCTACAA AAAACAAAGC
11101 AAAATAATAA CAGCAAAAGA AAAAAACATT TCATTGAGAT TCCGATTGAG
11151 ATTTGCATCA AATTACTTAG GTTATTTTGG GAGAATTAC ATCTTTATAG
11201 GATTGTTGGA TTTTCATATT TGAAATGATA AATCTCTCCA TTTTATTA
11251 TATTTTAACA TGTACCTCAG TAAAATTGTA TAGTTTCTT CAGTAAAGTT
11301 GTATAGTTTT CTTGCTATGA GTCTTACATT TTTATAAGGC TTTACTTTCAG
11351 ATGTTCTATC AGTTTTAAAA TGACCTGATT TTCTAAGTAG CAGGATAGTA
11401 TCCAGGTAAA GTAAACCCAC CTACCATACT TTTGGAAATA GGGGGATGAT
11451 GAAGATGACA AAGAATAGGA AGAAAGAGGA GGAGGGGGAG GAGGAGAAAA
11501 AGGAAAGGAA GAAGGGAAGA AGGAAGAAGA AGAAAGAAGA ACACAGCTAA
11551 AAGAATTTAC TAGGTTCTAG GCATTTCTCT AAGCCCTTTA CATGTAAATG
11601 TTTATTTAAT CTTTACCACA ACCCTATGAG ATAATTATCA TTCTCATTTT
11651 ACAGACGAAG AAACAGACGC ACTGAGAGTT TAAGTATATT CCCCCAAGGT
11701 CCCATAAGCA AAGATTGGAT TGGAAATCAG GGTGTTTGCC TCCAGAGCCT
11751 GTGTATTTTG TTCTCTTATG GCATGAGTGT ATTTGTAGGG ACACAGATTG
11801 AAAATGTTTT GACATTTATT GGAAGCATCA GGTTTTTTTC CTTCTGTTAC
11851 ACTACTAATC AATAAATGAG TTCTAATGTA AGGGAAAGCA TCGCCACACA
11901 GCTGGATGTA TGCTCTCACA TTCCCAGTTA CATAAGGTGC ATCAGCTCTT
11951 GAGGATGGGA CTGAGAATGG TTGAGAAAGA CAAGAGTCA CACTTCAAGA
12001 GTCTCCAATT CCAGTCTCTG AGATTCCAAC ACTCTACTTA AACTGAAGA
12051 ACTCAGAGCT GTGCTTCTT TTTGGGTTTAC ATGGGGGAAA TCTTAACCTT
12101 TCCTCCACTA AAAGTAAAAG ATTAAGTTGA CATCTCTATG GCCACCTTTT
12151 CCCTACATCA AGTGTTTTAA TAGGAACAGA AACTCCAGC TTTCTTTTG
12201 GATGAGTATT CCTCAGCCAT CCCACTTCTC TTGAGAGCAC TGGATTTTTC
12251 TTAGTAATCA GCATCCTTTG ACATAAAGGA AGAAAAGGAA AGGGCCACCT
12301 GTGTCATCTA TAGTTGAGGC TTAGGTAGGT TAGGGAGCAT CCTGGCTCTT
12351 TAGGGCCACT ACTCTAACAT ATGGTTCCAT GGATGTCATG GGTGAGGCAA
12401 CAGGGTTTGG AAATTTTGA ACTACTCTGC TGCAAACTCA GAGATTCCTA
12451 ACATATGGGG GTAATGAATT GACATTGCTG ATGACAAATA TAAGCAACTC
12501 TTGAGTATCT CAGTGAATTG AGAACTGAGG TACATAGATA TTCAGTGACT
12551 TCCAAAAGTT CCCATACAGC TGAACCAAGG ATTTCTTTCT TTCTTTCTTT

```

FIGURE 3, page 4 of 8

```

12601 CTTTCTTTCT TTCTTTCTTT CTTTCTTTCT TTCTTTCTTT CTTTCTTTCT
12651 TTCTTTCTTT TCTTTCTTTT TTTCTTTCTT TCTTTCTTTT TTTCTTTCTT
12701 CTTTCTCTTT CTTTCTCTCT TTTTCCTTTT CTTTTCTTTT CTTTTCTTTC
12751 TCTCTTTCTT TCTCTCTCTC TCTTTCTGTC TTCCTCCCTT CCTCCCTTTC
12801 TCTCTTTCTC CTTTCTTTTC TTCCCTTCCT CTTTACAGGC ATGCACCACC
12851 ATGCCCAGCT AATTTTGTGA TTTTGTAGTAG AGTACCGGGT TTCACCATGT
12901 TGGTCAGGCT GGTCTTGAAC TCCTGACCTC AGGTGATCCA CCCACCTCAG
12951 CCCCAGAAAG TGCTTGGGAT TACAGGTGTG AGCCACCGTG TGCACGGCTG
13001 GAACCAAGGA TTTCTAATTA GTTTTATTTT TTATTTTTTT TCTTTTGTAG
13051 AAGGAGTCTC ACTCTGTCTC CCAGGCTGGA GTGCAGTGGG GCAATCTCAA
13101 CTCACTGCAA CCTCTGCCTC GTGGGTCAA GTGATTCTCG TGCCTCAGCC
13151 TCCTGAGCAG CTGGGATTAC AGGCATGCCA TCATGCCTGG CTAATTTTTT
13201 TTTCTTTCTT TTTGAGACAG AGTCTCACTC TGTTGCCAG GCTGGATCGC
13251 AGTGGTGCAA TCACGGCCCA CTATTACCTC TGCCTCCAG GTTCAAGTAA
13301 TTCTCCTGCC TCAGCATCCC AGGTAGCTGG GAATACAGGT GCACGCCACC
13351 ACGCCTGACT AATATTTGTA TTTTGTAGCG AGATGGGGTT TCATCATGTT
13401 GGCCAGGCTG GTCTCGAACT CCTGACTTCA GGTGATCCAT CCGCCTGGC
13451 CTCCCAATGT GCTGGGATTA CAGGCATGAG TCACCGCGCC CAGCCTAACT
13501 AGGTATTTTA TGCACCTCTC CTAATCTCAG AAGTCTTCAT TAATTCACA
13551 AACATTTAT TACCTTTTATA TATGTTCCAG GTAATATGTT AGGCTATGGG
13601 AATACAGCAG TGAAGAAAAC ATGGTCCCCT CTGCCCTCAT GGAATTTTCA
13651 ATACACATTT TGACACATCA CTGAAGCTAA GTGTTCTAGA AACACACAAA
13701 CAATGTTAGT TCCTTGAACA AGATATACAT CAAAGAAGGG ACTTCTATTA
13751 GCAAGAGCGT TCTCTATGAG TCTCCTAAGA CTGGATTTT TCAGATAGAG
13801 TTCTTTCCGC CTTATTCAAT GTTTGCTCCG AAGCCTGCTT CATCAGCAA
13851 GTCTGCCTGA TACCTTTTATA TGTACTCTTC TCACGTTAGT GACTTCTCAA
13901 TGTTCTAAGA CCCATGCTTT TTAAGGAAGT TTATTTTGTA TATTTATATG
13951 ATTATTAAG TGTTACAGTA TATGTTTCATC ATGAGAAATT TAGAAAATAG
14001 AGAAATGTAG AGAAAAAGAT TTCTAAAAC GATATAAGAC TATCACACAC
14051 AAAAAAGAT ATTTTGGTTC ATTTTTCCTT TTTTGTGTC ATCTATTTTG
14101 TTTTATTGTA TATATTCAAG GTGTACAATG TGATGTTTCG ATGTATGTAC
14151 ACATTGTGAA ATGATTACCA CAACCAAAC AATTAACACA TTCATCACCT
14201 CACATAGTTA TCATTTTGTG ACGTGTGTGT GTGTGTGTGT GTGTGTGTGT
14251 GGTAAGCTT AAGATCTACT CTCTTTAAAA ATTTCAAGTA CACAATACAT
14301 TATTGTCAAC TATAGTCATC ATGTTGTACA TTAGAGCTCT GAACTTATT
14351 TATCTTATAA CTCTAAATTT GTAGCCTTTG ATCAAAATCC TTCTATTTC
14401 CTAAATCCCC ATCCCCTGGT AACCACCCAT TCTACTCTGT TGCTAGGTGT
14451 TCAACTTTT TAGATTCCAC ATATAAGTAA GACAATGCAG TATTTTCTT
14501 TATGTGTCTA GCTCATTTCA CTTAGCATAA TGTCTCTAG GTTCATCTGT
14551 GTTGTACAGG ATGGCAGAGT TTCTGTAAAT TTATGGTTGA ATAATATTCA
14601 TACACACACA CACACACACG CACACACACA CACACACACA CAGACACACC
14651 CACCAGATTT TCTTTATCCA TTCATCTGTC AACAGATACT GAGTTTGTGT
14701 CCATATCTTG GCTATTGAGA ATAATACTAC AATGAGCATG AGAGTGCAGA
14751 TATCTCTTTG AGATACTGAT TTCCTTTAGG TATACACCCA GCAGTGGGAT
14801 TATTTGATCG TTTGGCCGTT CTGTTGTAA TTTTGTGGA GAACCTCCAT
14851 GCTGTTTTTC ATAATGGCTG TGTCAAGTTA TGTTCACACA AACAGTGTAC
14901 AAGGTCTTTT TTCTTACATC CCCACCAACA CTTTTTTTTT TTAATAATAG
14951 CCATTCTAAC AGGTGTGAGG TGATATCTCA TTGTGGCTTT GATTGTCATT
15001 TTTGTGATGA TTAGTGATGT TGAACACCTT TTCATATACC TGTGGCCGT
15051 TTGTATACCG CCTTCGGAGA AAGTCTATTC AAGTGCATGC TATTGTGTTA
15101 CATAGCTGTG ATCATATTTG CATTGCTCTA TAACTGGAGC TCTCAAGTCT
15151 CACCCGTCAT CTCTCTGGAC CTCTGGGTTA TAAGTACAGC CTTTATTACC
15201 AACATTGACT GATTGCCTGT TTTTGTTTTT GTTTTGTGTT TTAACAGTTG
15251 TGCTAATGGT ATTTTCCCTG GTATCTGTCA CCTATGGGGC CACCTTTTGC
15301 AATATGTTGG CTATCCAGAT CAAGTACGAT GACTACAAGA TTCGCCTTGG
15351 GCCACTAGAA GTCCTCTGCA TCACCATCTG GCGGACATTG GAGATCACTT
15401 CCCGCCCTCT GATTCTGGTG CTCTTCTCAG CCACTTTGAA ATTGAAGGCT
15451 TGCCCTTCC TAGTCTCAA CTTCTGATC ATCTCTTTG AGCCTGGAT
15501 TAAGTTCTGG AGAAGTGGTG CCCAGATGCC CAATAACATT GAGAAAAACT
15551 TCAGCCGGGT CGGCACTCTG GTGGTCTGTA TTTCAGTCAC CATCCTCTAT
15601 GCTGGCATCA ACTTCTCTTG CTGGTCAGCT TTGCAGTTGA GGTTGGCAGA
15651 CAGAGATCTC GTCGACAAAG GGCAGAACTG GGGACATATG GGCCTGCACT
15701 ATAGTGTGAG GTTGGTAGAG AATGTGATCA TGGTCTGGT TTTTAAGTTC

```

FIGURE 3, page 5 of 8

```

15751 TTTGGAGTGA AAGTGTACT GAATTACTGT CATTCCTTGA TTGCCTTGCA
15801 GCTCATTATT GCTTATCTGA TTTCCATTGG CTTTCATGCTC CTTTTCTTCC
15851 AGTACTTGCA TCCATTGCGC TCACTCTTCA CCCATAATGT AGTAGACTAC
15901 CTCCATTGTG TCTGCTGTCA CCAGCACCCCT CGGACCAGGG TTGAGAACTC
15951 AGAGCCACCC TTTGAGACTG AAGCAAGGCA AAGTGTGTG TGATTCTATT
16001 TTCTGGGTAT TTTAGGAAGA GTTGGGAGTT GCCAAGAGTA ACCATGAAAT
16051 TGAACGAAAG GATGAGGTTC ATGGGTGAGA TACCCATCAG TACATTTTCT
16101 TGACTTTTCT GTTAAGCCTA TCAGAAGAAA GAGCAACTCC CAAATAGGTT
16151 TTATTTTCTT AAGAGTTACC ACTATGTTTG GAAACAGGGG GTATCGACTA
16201 TATAGTTGAA AGGGTCAGAA ATACCATTCA CACCCTTCTT ACCCAAGTCA
16251 ATTGGAATAA CTTGTCTTCA AACACTTTAG GCTCTCTAAA GTGACCTTCT
16301 AGCTCTGCTC ATTTGCTTGA TGCATTCTG AGCTTTCTCTG GGCTGAGCTG
16351 AAGGCCGAGA ATCCCGCTAG AATATATCCT GACTGATCAG AGGATATGAC
16401 AGCTTACCAG CTAAGAGTAC CTCCCAGGAA ACAGTCTGAC TAATGTGGAA
16451 CCTGCAACTG TCAGTGTGGC TGGGGTCTTT TTAATTCCAG TGAGAAGCTC
16501 TGGCTGAGAA GAAAATCACC ACTATTAAAA AAGCTGCTCC CCAAGCAGAT
16551 TAGCTCTCTG TTAGGATTTT ACTAGTGGCC ATTCAGCAAG GACCTCTCTT
16601 TACAGTGGCA CTTCATAGGC AACTCTAAG GAGAAAGTGC AGAGTAGAAT
16651 TCCTTCAGGG CATAAGCCAA AATGACTCTT TTTCTCAGGG ACCTGCATGG
16701 GCCTCCAGCT TGTCTATTGG AATTGTTAAG TGAAGCCTCT CACTTAGTGC
16751 CTCATTAGCA GAGATTTCTT CCAACCCAGC TTTTCTGTGC TCTTGGTATT
16801 TTAATACTTG ATGTGGACCT CAGAGAAGCT GAACTGTAAT TGAAAAATGTT
16851 TCCGATGTGT GGAAGAAATG AAGACTGCTT TGTGTCTGCT GTTGTCTTGA
16901 GTATTTTCAT AATGTGTGTG TGTGTGTGTG TGTGTGTGTG TGTGTATGTG
16951 TATGTGTGTA GGAAGAAAG TAATAATGGC TGAGACATCA CCTTCATGTT
17001 GTTTGCGATT GGGATGGGTG ACTAACACTC CAAGGTAGAG TGAAGGCAGA
17051 GGAGGGAAAC AAGATCACAT TAAATCATCA TCAGTACTGG TTTCTGCCTA
17101 CAGGAGTTTA CTTTTTTTTT TTTTCTTTT TTGAGATGGA GTCTCGCTCT
17151 GTTTTCTAGG CTGAAGTGCA GTGGTGTGAT CTTGGCTCAC TGCAGCCTCT
17201 GCCTCCTGGG TTCAAGCAGN NNNNNNNNNN NNNNNNNNNN NNNNAGTGAT
17251 CCACCCGCTT CCGTCTCCCA AAGCACTGGG ATTACAGGCA TGAGCCACCT
17301 CAGCGCGCCA GAGTTTACT TTATAACAAG GAACATATGT TTATCAACCC
17351 TCTGTTTCGT CCTATACCCC CAGTGGACGA ATGCATGTCT CCTTTTCTCC
17401 TATATCTCAA TGTTTACATC TCATATCAGT TGGGTATTTT GATAGGAATG
17451 TCAGCCAGCT ACCTCTGAGG TAACCAAGGG ATTGAAGTTA CTATGGCCAC
17501 TGCCTATTGG GACCAAATAT CCCAGCATTT ACCTAACTAA TGCTTGCCCC
17551 TCACAGACCA GGAAATTTAA AAGAACTCCT AGTCGTGGCC ACCACAACAC
17601 TTCAAGAAAT TGTGAACAAT CTGACCTAGG GCTTCCTGTC CTCATCCAAT
17651 TTTACTCTTG GTAGCATGCT AAGAATTTAT CTTTAGTCAT TTCCTCTCCT
17701 CTTATCCAAT GTCAGGACAT TATGTTGAGG GAGTTCTCTC TTCTAAGTAG
17751 CAGGGCTGTT AACCAAAGTA TCTTATTCTT TGGCATGGCT AGCATGGTTT
17801 TCCCTTCATC AGCCACTGTT TGGGACTAAA AGGATTATAT ACTTAATTTG
17851 GGAGAGACTG TATGGACTTG CTTTGGAACA GTGGAGAGCT CCTTCTTCA
17901 ACCCCAATC CCCCATTTCA TTTTTCATGA TGAAGAGACT TAGTTATTGT
17951 CATATAAAGC TCACCTGCTG TCTTCTAACT ATGTTATTCA AGG

```

FEATURES:

Gene Structure

FORM 1:

```

Start: 2001
Exon: 2001-2335
Intron: 2336-7242
Exon: 7243-7511
Intron: 7512-15247
Exon: 15248-15990
Stop: 15991

```

FORM 2:

```

Start: 1962
Exon: 1962-2335
Intron: 2336-7242
Exon: 7243-7511
Intron: 7512-15247

```

Exon: 15248-15993
Stop: 15991

CHROMOSOME MAP POSITION:
Chromosome 23

ALLELIC VARIANTS (SNPs):

DNA			
Position	Major	Minor	Domain
2584	G	C	Intron
2655	A	T	Intron
3693	G	A	Intron
3992	G	C	Intron
6285	-	A	Intron
7066	A	T	Intron
14223	-	T G	Intron
16915	-	G T	Beyond ORF (3')

Context:

DNA

Position

2584	ATAAACCGCTATCATTATTTATGCATCTAATCCTCTTGGGACCTGTTATCAGGTGAGCAA CTTTTAAATCTTTTCCTTACCCCCCTAACCCACCCACCCAGACTTGGGCAGAGAAAGATGAA AGATTTACAAGATGGATACTATGGCTCTAATCAATTCTCTCATTTCCCTCCCACTCTCGGC TTCCCTGTCTACCATTAGAAAACCTTACCTGAAATCTTAAATGCCACCATGATGAACATG TGGTATGTACTTGTGTTCCAAAACAATGAACGATGCTATTTGGGCTGTGTAACTAGAAT [G, C] GGAAACAACAAGACGTGATCACCTGTGCATGAAGGCCATAGCTGCAGAGTGTGTAATTTT ATTTAAAAAATTTTTTTTCTGAGACAAGGTCTTGCTCTGCCTCCCAAGGTACAGTGCA GTGGTGCGATCATGGCTCACTGCAGCCTTGATCTCCTGGGATCAAGCGAACCTCCACCT CAGCCTCCAAGTAGCTGGGACCAAGGAATGTGTACCATGCCTGGTTAATTAAAAA ATTTTATAGGCCGGGTGTGGTGGCTCATGCCTGTAATCCCAGCACTTTGGGAGGCTGAG
2655	TTTCCTTACCCCCCTAACCCACCCAGACTTGGGCAGAGAAAGATGAAAGATTTACAAG ATGGATACTATGGCTCTAATCAATTCTCTCATTTCCCTCCCACTCTCGGCTTCCCTGTCTA CCATTAGAAAACCTTACCTGAAATCTTAAATGCCACCATGATGAACATGTGGTATGTACT TGTGTTCCAAAACAATGAACGATGCTATTTGGGCTGTGTAACTAGAATGGGAACAACAA GACGTGATCACCTGTGCATGAAGGCCATAGCTGCAGAGTGTGTAATTTTATTTAAAAA [A, T] TTTTTTTTTCTGAGACAAGGTCTTGCTCTGCCTCCCAAGGTACAGTGCAAGTGGTGCGATC ATGGCTCACTGCAGCCTTGATCTCCTGGGATCAAGCGAACCTCCCACTCAGCCTCCAAG TAGCTGGGACCAAGGAATGTGTACCATGCCTGGTTAATTAAAAAATTTTATAGG CCGGGTGTGGTGGCTCATGCCTGTAATCCCAGCACTTTGGGAGGCTGAGGCCGGGTGGATC ACCTGAGGTCAGGAGTTCAAGACCAGCTGGCCACATGGTGAAACCCCTGTCTCTACTAA
3693	TCTCTGTCGGTTTATCAGTTTCCTATTTATCTCTTTGTATATTTCTGCAATAAAGATACG AAGTTGGGAGGGGGCAAAGGAAGGCAGTTTCATCTCTATGTGGATGCAGTAGCACAAATT TAATAGTATCAAGTATTTCCATTAGATTGCCTTGAAGTGGAAAGAATGCACTTAATCCT AGCGAGATAGGCACCTGTGTCAACAGTCTCATCTGGATGCTATGGGGTTTTCAAGGTAGA GAGATGTTGCAAACTTATGAGTTCAGGAGTAAGGAATGGACCAAGTTTGTCTTGATTGC [G, A] AGAGAGGCAGACAACCTGCAGTCAGCCGAGGAATATGGGTGAGAGTGTGCAATGGGAAGA TACCTCATCATTAGACAACCTAAAAAGTCTGTGAACTAATTAAGGATGGAACTCACTCCT TTATAAAATTTTATATCTGTACACATGTATAATTTTATTTGCTACTTATACCTCAATAA GGCCAAAAAATTTTATCAATAAATTTTAAAGTGGGGAGGAATCGATTAGGCTCTATC AGAGAGAATATGGGATATCAATGGAAACAGTGGCCTGAAATTTGGAGTCTAGTCTTCCGC
3992	CGAGAGAGGCAGACAACCTGCAGTCAGCCGAGGAATATGGGTGAGAGTGTGCAATGGGAA GATACCTCATCATTAGACAACCTAAAAAGTCTGTGAACTAATTAAGGATGGAACTCACTC CTTTATAAAATTTTATATCTGTACACATGTATAATTTTATTTGCTACTTATACCTCAAT AAGGCCAAAAAATTTTATCAATAAATTTTAAAGTGGGGAGGAATCGATTAGGCTCTA

